BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

C4

XP-002105952 PD 1986
P 1434-48 = (15)

Volume 14 Number 8 1986 Nucleic Acids Research

DNA sequence of the region in the genome of herpes simplex virus type 1 containing the exonuclease gene and neighbouring genes

Duncan J.McGeoch*, Aidan Dolan and Margaret C.Frame

MRC Virology Unit, Institute of Virology, University of Glasgow, Church Street, Glasgow G11 5JR, UK

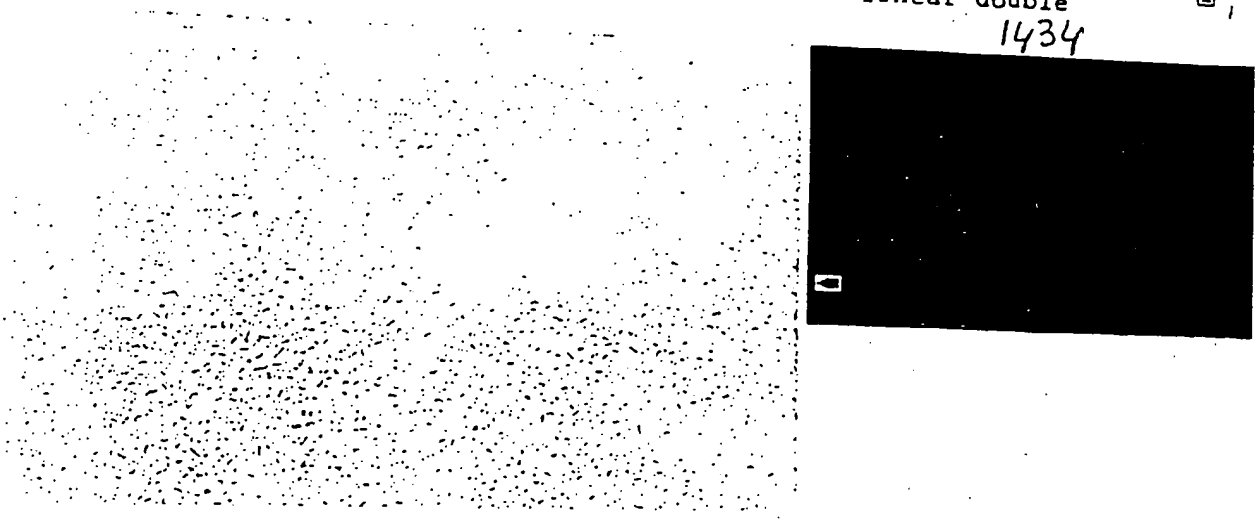
Received 21 February 1986; Accepted 19 March 1986

ABSTRACT

We report the sequence of a 7800 base pair region of herpes simplex virus type 1 DNA, representing approximately 0.16 to 0.20 map units in the genome. This contains sequences transcribed into a leftward oriented set of five 3' coterminal mRNAs, together with two rightward transcribed flanking genes. One of the leftward genes encodes the virus's alkaline exonuclease, but the other gene products are uncharacterized. The amino acid sequence of one encoded protein suggested that it is a membrane embedded species. The DNA sequence is densely utilised, with two predicted out-of-frame overlaps of coding sequences, and probably six occurrences of promoter elements within coding sequences. Homologues of five of the genes were found for the distantly related Epstein-Barr virus, with a similar overall relative arrangement.

INTRODUCTION

The genome of herpes simplex virus (HSV) is a linear double



1434

INTRODUCTION

The genome of herpes simplex virus (HSV) is a linear double stranded DNA of about 155,000 base pairs (1) and is thought to possess about 80 genes. The genome contains several genes specifying proteins involved in its replication, as well as sequences which act as origins of DNA synthesis, and others which are signals or sites in cutting and packaging of nascent DNA (2,3). The HSV proteins which participate in DNA replication include enzymes involved in the metabolism of DNA precursors (for instance, thymidine kinase and ribonucleotide reductase (4,5)), and also species more directly concerned with DNA synthesis. The latter include a DNA polymerase, an exonuclease and a polypeptide designated operationally as the "major DNA binding protein" (6,7,8). As part of a large scale sequence analysis of HSV-1 DNA, we recently reported the sequences of the genes for DNA polymerase and the major DNA

binding protein (9). In this paper, we describe the DNA sequence of the exonuclease gene and of the neighbouring genes.

The exonuclease was first detected as an increased alkaline nuclease activity present in HSV infected cells (10) and was subsequently purified and shown to be a single polypeptide chain (7,11,12,13), possessing both 5' and 3' exonuclease activity and also an endonuclease activity (11,14,15). Analysis of temperature sensitive mutants of the HSV-2 exonuclease gene has shown that the enzyme is necessary for efficient replication of virus DNA, although its precise role has not been defined (16,17,18). The exonuclease gene was located in the long unique region (U_L: see Figure 1) of HSV-1 DNA, near 0.17 map units, by hybrid arrest of translation of active enzyme (19). Analysis of the transcript organization of this region of the genome has shown that exonuclease mRNA is a member of a 3' coterminal set of mRNAS (20). This paper reports a 7800 bp sequence containing the whole of the 3' coterminal family, for HSV-1 strain 17, together with all or part of the two flanking genes.

MATERIALS AND METHODS

(1) Plasmids.

Cloned restriction nuclease fragments of HSV-1 strain 17 DNA were used for sequence analysis, as follows: EcoRI g and EcoRI d in the EcoRI site of pACYC184, from V.G. Preston; and KpnI f and KpnI g cloned into the PstI site of pAT153 by dG/dC tailing, from A.J. Davison.

(2) Sequence analysis and interpretation.

DNA sequences were determined by M13/dideoxy methods, as described (9,21,22). Computing was performed with a DEC PDP 11/44 under RSX11M, with programs as previously described (9,22).

(3) Description of the HSV-1 genome.

HSV genetic and mRNA mapping data commonly specify genome locations in terms of fractional map units, running from 0.000 to 1.000. There exist significant discrepancies between the numbering for HSV-1 strain 17 used by us and, for instance, strain KOS (20,23); these are of the order of 0.005 map units or 800 bp. We deal with this, for the present, by quoting map units as approximations to two decimal places only.

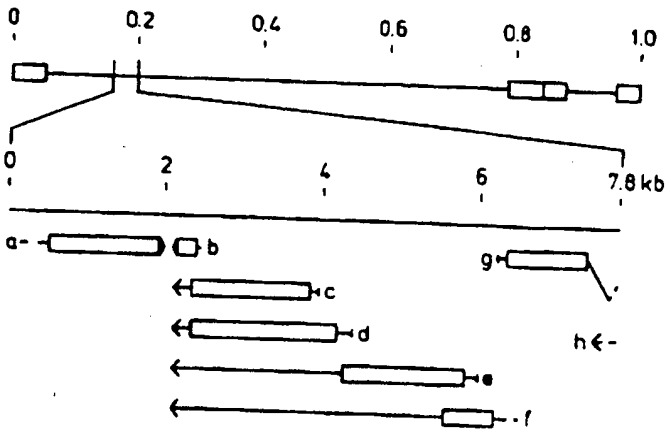


Figure 1. Transcript arrangement of the HSV-1 genome, 0.16 to 0.20 map units. The upper part of the figure shows a conventional representation of the HSV-1 genome, with major repeat elements as open boxes, and the long and short unique sequences (U_L and U_S) as solid lines. The scale indicates fractional genome length map units. In the lower part of the figure an expansion is given for the region 0.16 to 0.20 map units, with numbering in kb corresponding to Figure 2. Locations and orientations of mRNAs are shown as arrows (20,23,24), with predicted coding regions as open boxes. Note that transcript b is hypothetical (see text). For transcript g, only the first exon and part of the intron region are included (23).

RESULTS

(1) Organization of the HSV-1 genome, 0.16 to 0.20 map units.

E. Wagner's laboratory has mapped mRNA species transcribed from HSV-1 DNA in the region 0.16 to 0.20 map units (20,23,24). Figure 1 indicates the location of this region in U_L, and presents a transcript map based on their studies and on our sequence data. It presently appears that all or part of at least eight mRNA species originate from the region. For this paper we have designated these a to h (Figure 1). RNAs b, c, d, e and f form a leftward transcribed 3' coterminal family. The two longest of these, e and f, were described as late transcripts of 3.9 and 4.5 kb (20). d is an early 2.3 kb transcript expressing the exonuclease (20). c is a late 1.9 kb species (20). b is a species suggested by our sequence data and is still hypothetical. These RNAs share a common polyadenylation site. To the left of this 3' coterminal group, RNA a is a 2 kb late, rightward transcribed species (24). To

the right of the group, RNA g is a late, rightward 2.7 kb transcript, with a large intron (23). Only the first exon is in the region treated here. Within the intron region are 3' coterminal, leftward transcripts (23); the downstream termini of these are shown generically as h in Figure 1.

Organization in 3' coterminal families is a common feature of HSV gene arrangement (22,24). It is thought that each RNA is translated to give the polypeptide encoded adjacent to the 5' terminus, while distal reading frames remain unused. Figure 1 also shows for mRNA species a to g the locations of the proposed, corresponding protein coding regions, as deduced from our DNA sequence data. These open reading frames were evaluated using published mRNA mapping data (20,23,24), by analysis of codon usage compared with known HSV genes (25), and by comparisons with corresponding sequences from the genomes of varicella-zoster virus (VZV) and Epstein-Barr virus (EBV). VZV is, like HSV, a member of the alphaherpesvirinae sub-family (26), although its DNA sequence and many details of genome organization differ substantially from those of HSV. The unpublished sequence of the corresponding region of the VZV genome was made available to us by our colleague A.J. Davison. EBV is a member of the gammaherpesvirinae, and its complete genome sequence has been published (26,27).

We have determined the DNA sequence of HSV-1 strain 17 for the region shown in Figure 1. This sequence is presented in Figure 2 as 7800 bp, of base composition 65.2% G+C. Proposed encoded amino acid sequences are also shown. The sequence was determined by the M13/dideoxy system with random sub-fragments of four large, plasmid cloned fragments of HSV-1 DNA (KpnI f and g, and EcoRI d and g). KpnI f and g lie to the left and right, respectively, of the KpnI site at residue 5958 in Figure 2, and EcoRI d and g lie to the left and right, respectively, of the EcoRI site at 6670. The sequence as presented starts at an arbitrary point to the left of the mRNA a region and ends downstream of the mRNA g splice donor site. In the following sections each gene is treated in turn.

(2) Gene "a" encodes a hydrophobic protein.

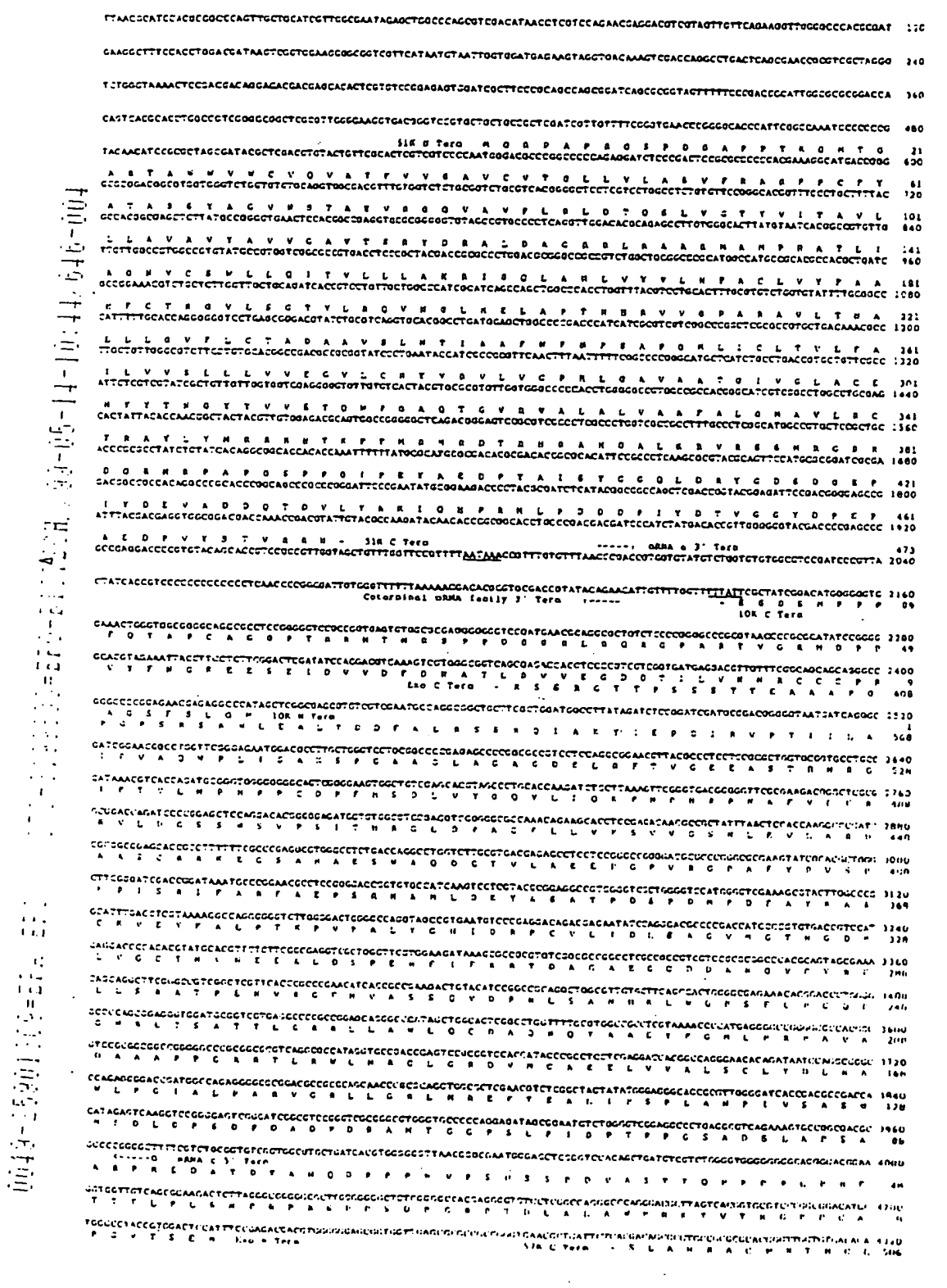
A rightward transcribed 2 kb mRNA has been mapped to the

left of the exonuclease gene region (24). We consider that the 3' terminus of this RNA is near residue 2002 of Figure 2, downstream of the appropriately placed polyadenylation consensus AATAAA at 1977. The 5' terminus should therefore be near the start of the sequence presented here, but has not been mapped precisely. We propose that the protein coding region of gene a starts with ATG at residue 538 and closes with TAG at 1957. This reading frame of 473 codons encodes a protein of M_r 51389, now termed 51K.

No known protein or function has been assigned clearly to this gene. The encoded protein contains a high content of hydrophobic amino acids (the three most abundant amino acid species are Ala, Val and Leu), and also an excess of basic over acidic residues. The hydrophobic residues are notably clustered. There are at least five regions which, from their degree of hydrophobicity and absence of charged residues, could span a lipid bilayer membrane. We think it possible that 51K is a previously undescribed, integral membrane protein, similar to that encoded by HSV-1 near 0.74 map units which is involved in virus-induced cell fusion (28). The 51K amino acid sequence possesses low but definite homology with that of the EBV reading frame BBRF3 (27) (see section 8, below).

(3) A possible small gene at the 3' end of the coterminal family.

As described in the next section, mRNAs c, d, e and f have their 3' termini near residue 2115. Upstream of this (on the leftward 5'- 3' strand) lies the exonuclease coding region, which terminates at TGA, residue 2343. There also exists a small open reading frame overlapping the downstream end of the exonuclease reading frame, out of phase, from ATG at 2425 to TAA at 2137. This would encode a protein of 96 amino acids, M_r 10486, now called 10K. There is no mapped mRNA corresponding to this reading frame. However, we think this small gene is probably real, for the following reasons. Both EBV and VZV possess corresponding open reading frames, with amino acid sequence homology to the HSV-1 candidate. For the EBV example (reading frame BBLF1), a promoter has been identified (27). Lastly, the 10K reading frame shows a reasonable codon usage. A



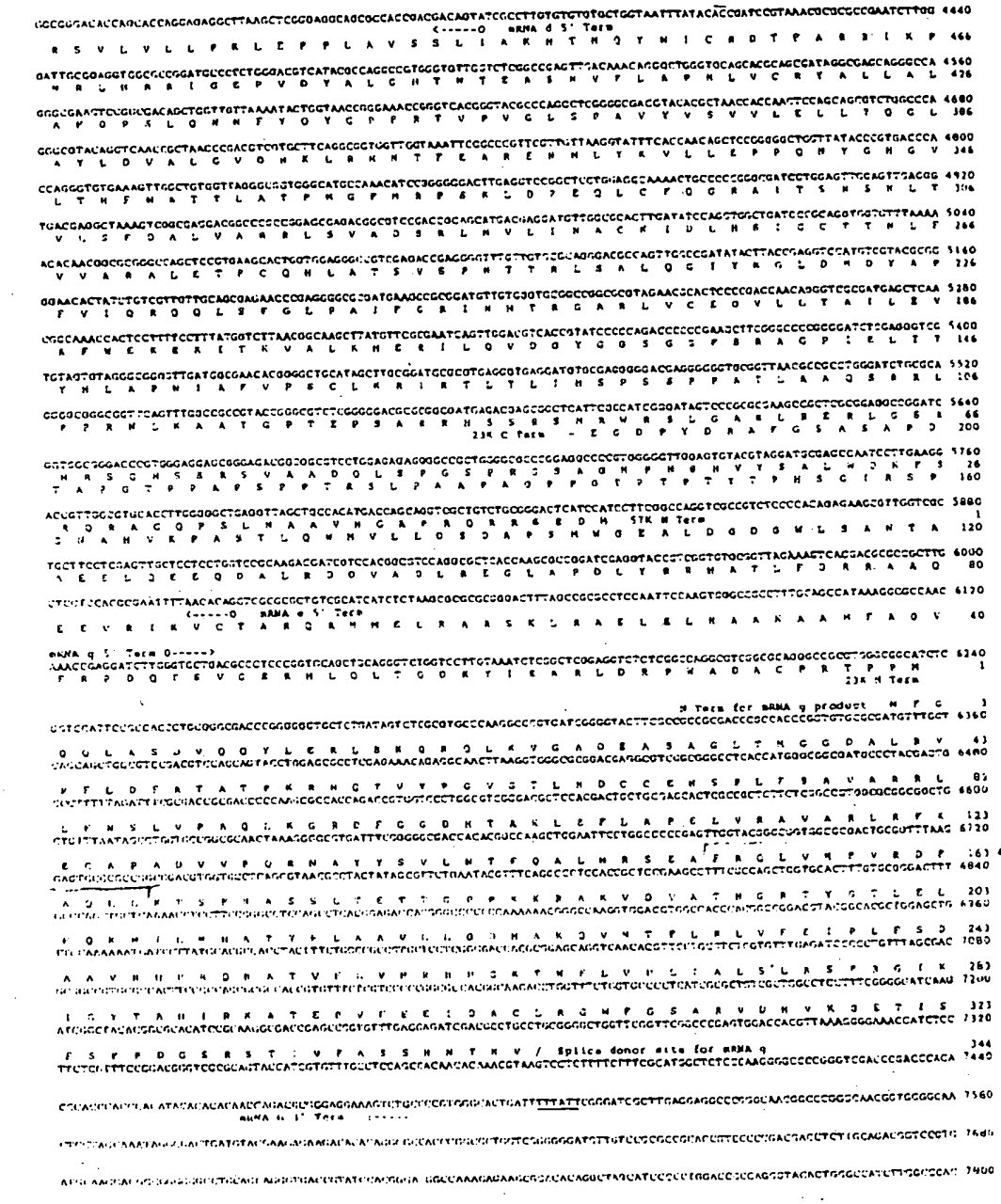


Figure 2. DNA sequence of the HSV-1 genome, 0.16 to 0.20 map units. The DNA sequence is shown for the expanded region of Figure 1, as the rightward 5'-3' strand only. Precisely mapped 5' termini of mRNAs are shown as O-----> (20,23), and presumed 3' termini as ------:. Candidate polyadenylation sequences AATAAA are underlined. Predicted amino acid sequences are given in single letter code, with those from rightward 5'-3' transcripts above the DNA sequence, and from leftward transcripts below.

hypothetical mRNA species (b) for this protein is indicated in Figure 1.

(4) The exonuclease gene.

Costa et al. (20) have determined a 1000 bp sequence for HSV-1 strain KOS, corresponding to residues 3656 - 4654 of Figure 2. An extra residue is present at position 6 in the KOS sequence (presumed an error), and there are 10 base substitutions. Within this sequence Costa et al. mapped the 5' termini of the mRNAs here called c and d. mRNA d encodes the exonuclease and has its 5' terminus at position 4380 of Figure 2 (20). From the estimated size of RNA d (2.3 kb) it is clear that the common 3' termini of this mRNA family lie near residue 2115, downstream of the polyadenylation consensus AATAAA at 2134 on the leftward 5'-3' strand.

We consider the exonuclease coding region runs from ATG at 4221 to TGA at 2343. This gives a protein of 626 amino acids, M_r 67503, which corresponds moderately with a recent estimate of the M_r from gel electrophoresis, of 85000 (13). The origin and function of mRNA c, whose 5' terminus was mapped to position 3971 (20), are less clear. There is no obvious TATA box upstream of RNA c's 5' terminus. Downstream of the 5' terminus, the first two potential initiator ATG codons are at 3843 and, 174 nucleotides later, at 3669; both of these are in the exonuclease assigned reading frame. It is therefore likely that translation of RNA c would give rise to an N-terminally truncated exonuclease protein. Starting from the first ATG, at 3843, the translation product would contain 500 amino acids, with M_r 54395.

We have found that the exonuclease amino acid sequence is clearly related to EBV reading frame BGLF5 (27) (see section 8, below). Interestingly, however, the EBV sequence lacks residues corresponding to approximately 110 amino acids at the N terminus of HSV-1 exonuclease: that is, the EBV sequence is roughly equivalent in extent to the 500 amino acid polypeptide proposed to be translated from RNA c, reinforcing the notion that this could be a functional protein.

(5) Upstream neighbour of the exonuclease gene.

In a limited sequence analysis, Costa et al. (23) located

the 5' terminus of mRNA e at a position corresponding to residue 6025 (Figure 2), on the leftward 5'-3' strand. Curiously, this is just upstream of the best nearby TATA box candidate, at 6013 to 6020. We think that translation of this mRNA starts with the first ATG, at 5836, and ends with TGA at 4282. This would give a protein of 518 amino acids, M_r 57193, now called 57K. This assignment places the promoter, 5' terminus and part of the 5' non coding region of the next downstream transcript, RNA d, within the protein coding region of RNA e. This arrangement has precedents in a number of other HSV genes (29,30). The 57K amino sequence is homologous to that of the EBV reading frame BGLF4.

Costa et al. (20) tentatively identified either RNA e or RNA f as encoding a nucleocapsid protein designated VP19C. However, they apparently confused VP19C with another species, VP18.8 (31,32), and we conclude that there is presently no firm evidence as to the identity of 57K or of the RNA f product, 23K, described next.

(6) The 5' member of the coterminal RNA family.

The largest member of the leftward reading 3' coterminal family, RNA f, has been reported to be a 4.5 kb transcript (20). The 5' terminus of RNA f has not been mapped precisely, but we think it probable that it lies near 6470 or near 6590, downstream of TATA candidate sequences at 6484 and 6604. The most likely protein coding region runs from ATG at 6237 to TGA at 5592, encoding a protein of 215 amino acids, M_r 23454, which is now termed 23K. This assignment of RNA f's coding region thus includes the promoter and 5' terminal region of RNA e, and 82 codons (out of phase) of the 57K coding sequence. Arrangement of coding regions is proposed to be similar in the corresponding region of the VZV genome (A.J. Davison, personal communication). However, our interpretation should be regarded as tentative, until these coding regions can be verified experimentally.

(7) First exon of spliced RNA "g".

A major recent surprise in analysis of HSV's gene organization was the description of a gene with a 4 kb intron, within which were located genes in the opposite orientation

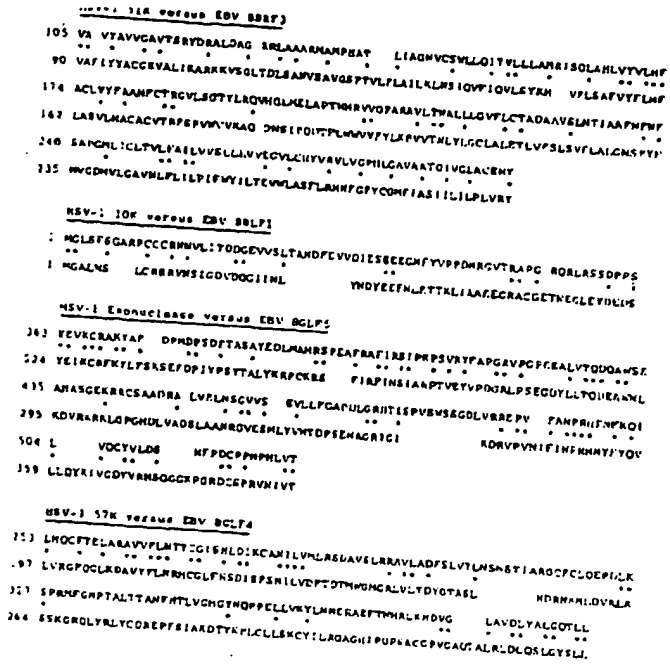


Figure 3. Alignments of HSV-1 and EBV predicted amino acid sequences. For four predicted HSV-1 amino acid sequences, alignments are shown with corresponding EBV sequences (27), for the most homologous regions only, as obtained with an alignment optimising program (35). In each case the HSV-1 sequence is the upper, and aligned identical residues are marked by asterisks.

(23). The spliced mRNA is that denoted as the rightward transcript g in Figure 1. The sequence of Figure 2 contains the upstream exon only. Partial sequence data were presented for this gene in HSV-1 strain KOS by Costa et al. (23). From their results, the 5' terminus of the mRNA lies at 6136 and the first exon ends with residue 7384 (Figure 2). Translation is thought to start with ATG at 6352. Within the intron region, the 3' termini of leftward transcribed mRNAs (denoted by h in Figure 1) are near 7482, downstream of AATAAA at 7510. These assignments place the promoters and 5' termini of the divergently transcribed RNAs f and g each within the other's protein coding region.

(8) Relations with genes of EBV.

The predicted amino acid sequences of the HSV-1 genes described above were compared with predicted amino acid sequences encoded in the EBV genome (27) by first using database

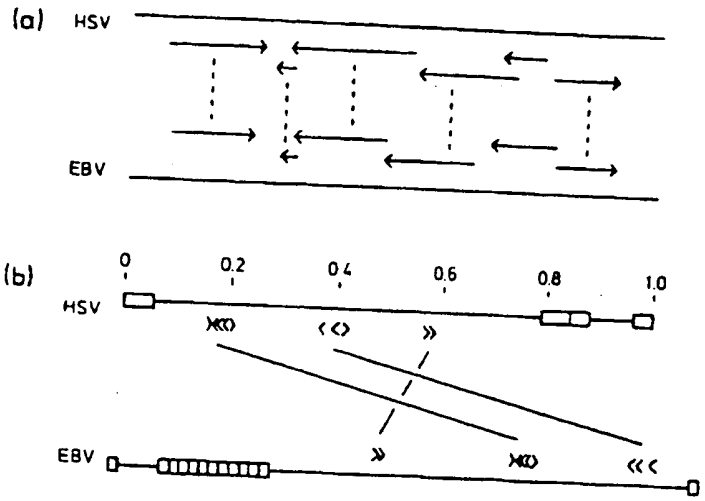


Figure 4. Relations between HSV-1 and EBV genome organizations. (a) The arrangement of reading frames from the sequence of Figure 2 is shown, and is compared with reading frames from the EBV sequence, residues 118501 to 126500 (27). Homologous pairs are indicated by dashed lines. For HSV-1 the reading frames, left to right, are for 51K, 10K, exonuclease, 57K, 23K and the first exon of the spliced gene. For EBV the reading frames are BBRF3, BBLF1, BGLF5, BGLF4, BGLF3 and BGRF1 (27). (b) The positions of 4 groups of homologous genes are compared in the two complete genomes. The position and orientation of each gene is shown as ">" or "<". The HSV-1 groups are: near 0.2 map units, genes discussed in this paper; near 0.4 map units, the gB, dbp, and pol genes (10); and near 0.6 map units, genes for ribonucleotide reductase (30,36).

searching methods (33), then "matrix" comparisons (34), and finally an optimal alignment program (35). These analyses showed that HSV-1 genes encoding 51K, 10K, exonuclease and 57K are homologous to EBV reading frames BBRF3, BBLF1, BGLF5 and BGLF4, respectively. The homologies are low (less, for instance, than that seen with HSV-1 and EBV DNA polymerases (9)) but are generally convincing. Figure 3 shows alignments of parts of the four pairs of sequences exhibiting the most conservation. In addition, the exons of the intron-containing HSV-1 gene (g in Figure 1) have previously been shown to be homologous to EBV's BGRF1 and BDRF1 (23). HSV-1 23K is presumed to correspond to EBV BGLF3, from their positions relative to the other reading frames, but no quantifiable amino acid sequence homology was detected.

Thus, despite extensive divergent evolution, the region of

the HSV-1 genome described in this paper, and the EBV genome region, residues 119000 to 126000, are recognizably preserved in gene layout and in encoded proteins, as shown in Figure 4(a). When these data are combined with homology relations for HSV and EBV ribonucleotide reductase proteins (36), and with our previous comparison of the DNA polymerase regions (9), the beginnings can be seen, in Figure 4(b), of an overall view of relations between the genome organizations of the two viruses.

DISCUSSION

Our interpretation of protein coding regions in this part of the HSV-1 genome, between 0.16 and 0.20 map units, indicates a dense use of the DNA sequence. There are two proposed, out-of-frame overlaps of coding sequences, of 27 and 82 codons, and the remaining gaps between adjacent coding sequences are minimal. In addition, since HSV promoters generally occupy 80-100 bp upstream from the cap site (37,38), it is clear that the promoters for RNAs of the 3' coterminal family, and also the promoter for RNA g, must lie within protein coding sequences. In the absence of direct information on amino acid sequences, the identification of coding regions is, of course, subject to errors of misinterpretation, and is also critically dependent on the absence of frameshifting errors in the DNA sequence. However, we regard our reading frame assignments as conservative, and note that they are generally supported by codon usage analysis and by comparison with evolutionarily related sequences from EBV and VZV.

Of the seven predicted sequences, only that for the exonuclease possesses any significant characterization. The predicted occurrence of an N-terminally truncated exonuclease fragment is intriguing, and the fact that EBV encodes a homologue of about the same extent as this shortened protein supports the notion that it could be functional. The 51K protein is the only one of the unknowns whose sequence gives any direct clue to its nature, in this case as a membrane-embedded species. Finally, from the mapping results of Lemaster and Roizman (31), one of our predicted species could be the virion protein VP18.8, but this remains unresolved.

The end result of our analyses is a clear view of gene organization in this portion of HSV-1 DNA; this has been greatly facilitated by the extensive mRNA mapping data of Costa et al. (20,23). Some qualifications remain, and these will only be resolved by extensive, further studies. We see the most immediate of these as being, first, an evaluation of the N-terminally truncated protein thought to be translated from RNA c, and second, direct analysis of our proposed RNA b.

Our comparisons with EBV showed that the part of the HSV-1 genome examined shows general colinearity of organization with a region of the EBV genome. Such colinearity also extends rightward, at least in that the two genomes possess correspondingly placed and homologous second exon sequences for the rightward, spliced RNA g (23). However, Figure 4(b) shows that larger scale rearrangements have certainly occurred during the divergent evolution of these viruses.

This sequence will be deposited with the EMBL Sequence Library.

ACKNOWLEDGEMENTS

We acknowledge the great help given us by A.J. Davison in discussion and access to unpublished data. We also thank C.M. Preston for criticism of the manuscript, and A.J. Davison and V.G. Preston for provision of plasmids.

*To whom correspondence should be addressed

REFERENCES

1. Kieff,E.D., Bachenheimer,S.L. and Roizman,B. (1971) J. Virol. 8, 125-132.
2. Stow,N.D. and McMonagle,E.C. (1983) Virology 130, 427-438.
3. Stow,N.D., McMonagle,E.C. and Davison,A.J. (1983) Nucl. Acids Res. 11, 8205-8220.
4. Kit,S. and Dubbs,D. (1963) Biochem. Biophys. Res. Comm. 11, 55-59.
5. Cohen,G.H. (1972) J. Virol. 9, 408-418.
6. Purifoy,D.J.M., Lewis,R.B. and Powell,K.L. (1977) Nature 269, 621-623.
7. Morrison,J.M. and Keir,H.M. (1968) J. Gen. Virol. 3, 337-347.
8. Powell,K.L., Littler,E. and Purifoy,D.J.M. (1981) J. Virol. 39, 894-902.

9. Quinn,J.P. and McGeoch,D.J. (1985) Nucl. Acids Res. 13, 8143-8163.
10. Keir,H.M. and Gold,E. (1963) Biochim. Biophys. Acta 72, 263-276.
11. Strobel,M. and Francke,B. (1980) Virology 103, 493-501.
12. Hoffmann,P.J. and Cheng,Y.-C. (1978) J. Biol. Chem. 253, 3557-3562.
13. Banks,L., Purifoy,D.J.M., Hurst,P.-F., Killington,R.A. and Powell,K.L. (1983) J. Gen. Virol. 64, 2249-2260.
14. Hoffmann,P.J. and Cheng,Y.-C. (1979) J. Virol. 32, 449-457.
15. Hoffmann,P.J. (1981) J. Virol. 38, 1005-1014.
16. Moss,H., Chartrand,P., Timbury,M.C. and Hay,J. (1979) J. Virol. 32, 140-146.
17. Francke,B. and Garrett,B. (1982) Virology 116, 116-127.
18. Moss,H. (1986) J. Gen. Virol., in press.
19. Preston,C.M. and Cordingley,M.G. (1982) J. Virol. 43, 386-394.
20. Costa,R.H., Draper,K.G., Banks,L., Powell,K.L., Cohen,G., Eisenberg,R. and Wagner,E.K. (1983) J. Virol. 48, 591-603.
21. Bankier,A.T. and Barrell,B.G. (1983) In Techniques in the Life Sciences, Flavell,R.A. Ed., Vol B508, pp.1-34, Elsevier, Ireland.
22. McGeoch,D.J., Dolan,A., Donald,S. and Rixon,F.J. (1985) J. Mol. Biol. 181, 1-13.
23. Costa,R.H., Draper,K.G., Kelly,T.J. and Wagner,E.K. (1985) J. Virol. 54, 317-328.
24. Wagner,E.K. (1985) In The Herpesviruses, Roizman,B. Ed., Vol 3, pp.45-104, Plenum Press, New York.
25. Staden,R. and McLachlan,A.D. (1982) Nucl. Acids Res. 10, 141-156.
26. Matthews,R.E.F. (1982) Intervirology 17, 1-199.
27. Baer,R., Bankier,A.T., Biggin,M.D., Deininger,P.L., Farrell,P.J., Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Seguin,C., Tuffnell,P.S. and Barrell,B.G. (1984) Nature 310, 207-211.
28. Debroy,C., Pedersen,N. and Person,S. (1985) Virology 145, 36-48.
29. Rixon,F.J. and McGeoch,D.J. (1984) Nucl. Acids Res. 12, 2473-2487.
30. McLauchlan,J. and Clements,J.B. (1983) J. Gen. Virol. 64, 997-1006.
31. Lemaster,S. and Roizman,B. (1980) J. Virol. 35, 798-811.
32. Braun,D.K., Batterson,W. and Roizman,B. (1984) J. Virol. 50, 645-648.
33. Devereaux,J., Haeberli,P. and Smithies,O. (1984) Nucl. Acids Res. 12, 387-395.
34. Pustell,J. and Kafatos,F.C. (1982) Nucl. Acids Res. 10, 4765-4782.
35. Taylor,P. (1984) Nucl. Acids Res. 12, 447-456.
36. Gibson, T., Stockwell,P., Ginsburg,M. and Barrell,B. (1984) Nucl. Acids Res. 12, 5087-5099.
37. McKnight,S.L. and Kingsbury,R. (1982) Science 217, 316-324.
38. Everett,R.D. (1983) Nucl. Acids Res. 11, 6647-6666.

Nucleic Acids Research, Vol. 18, No. 10 3065

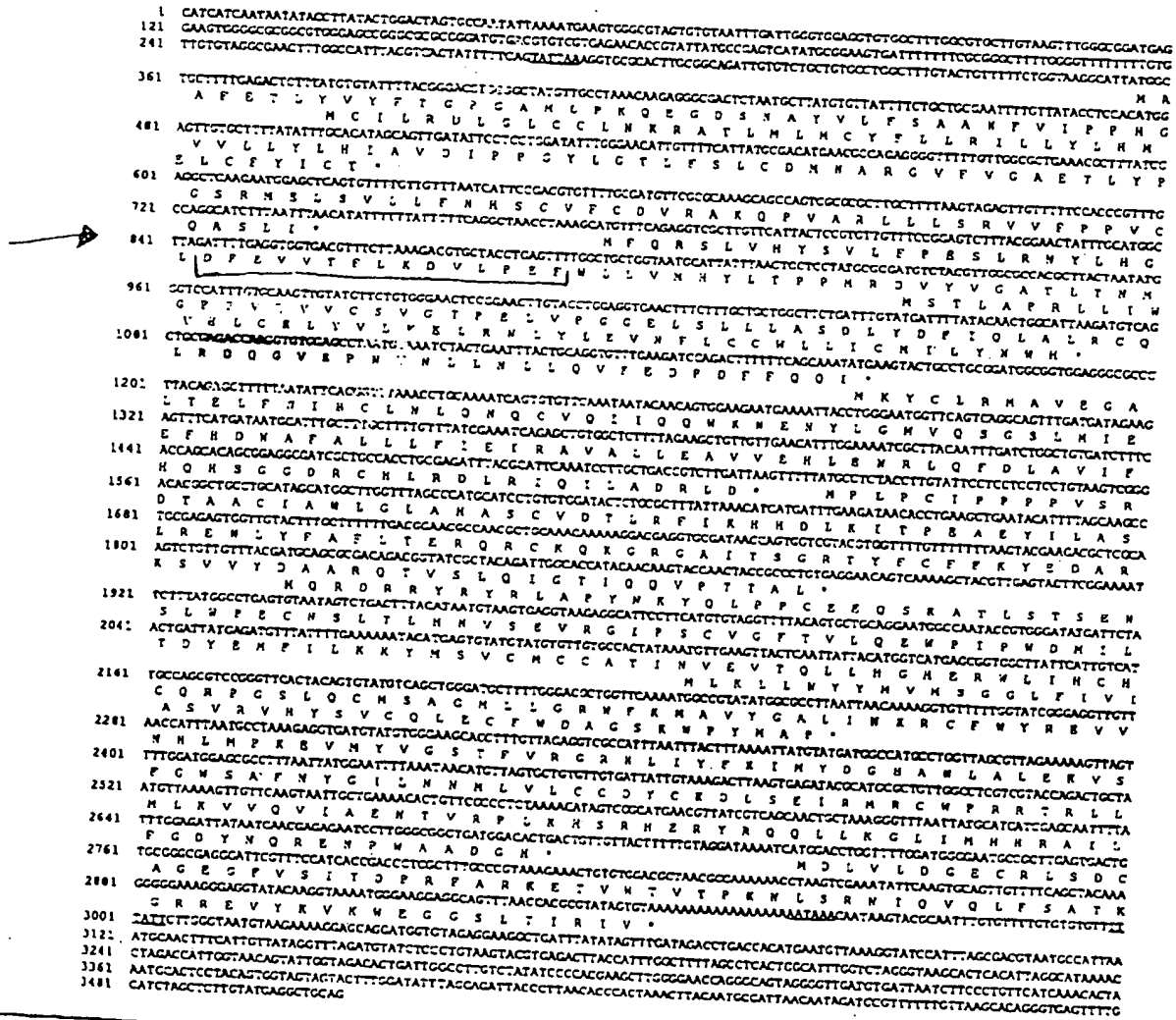
Nucleotide sequence of the right 10% of adenovirus type 12 DNA encoding the entire region E4

T.Hogenkamp and H.Esche*
Institute of Molecular Biology (Cancer Research), University of Essen, Hufelandstrasse 55, D-4300 Essen 1, FRG

Submitted April 2, 1990

EMBL accession no. X51800

We report the sequence of 3.5 kb of the right end of the adenovirus type 12 genome, containing the entire coding capacity of early region 4. The nucleotide sequence was determined by the Sanger dideoxy method (1). 189 bp of the right ITR of Ad12 published by Shinagawa and Padmanabhan (2) show 98% homology with the equivalent region of our nucleotide sequence and a high degree of homology with region E4 of the *in vivo* nononcogenic adenovirus types 2 and 5 (3, 4). The theoretical amino acid sequences corresponding to the various open reding frames are indicated. The putative TATA-box at nt 283 and the poly (A) addition site of region E4 at nt 2960 are underlined. The putative poly (A) addition site of the fiber mRNA transcribed



* To whom correspondence should be addressed

in opposite direction from the complementary DNA strand is located at nt 2998. Cloning of specific, complex spliced mRNAs transcribed from early region E4 using PCR and analysis of specific gene products are under investigation.

ACKNOWLEDGEMENTS

Supported by the Deutsche Forschungsgemeinschaft through SFB 102 and the Fonds der Chemischen Industrie.

REFERENCES

1. Sanger,F. (1981) *Science* **414**, 1205–1210.
2. Shinagawa,M. and Padmanabhan,R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3831–3835.
3. Hérissé,J., Rigolet,M., Dupont de Dinechin,S. and Galibert,F. (1981) *Nucl. Acids Res.* **9**, 4023–4042.
4. Steenbergh,P.H. and Sussenbach,J.S. (1979) *Gene* **6**, 307–318.